

SEQUENCE LISTING

(1) GENERAL INFORMATION:

- (i) APPLICANT: Darzins, Aldis
Mrachko, Gregory T.
- (ii) TITLE OF INVENTION: A Sphingomonas Biodesulfurization Catalyst
- (iii) NUMBER OF SEQUENCES: 13
- (iv) CORRESPONDENCE ADDRESS:
(A) ADDRESSEE: Hamilton, Brook, Smith & Reynolds, P.C.
(B) STREET: Two Militia Drive
(C) CITY: Lexington
(D) STATE: Massachusetts
(E) COUNTRY: USA
(F) ZIP: 02173
- (v) COMPUTER READABLE FORM:
(A) MEDIUM TYPE: Floppy disk
(B) COMPUTER: IBM PC compatible
(C) OPERATING SYSTEM: PC-DOS/MS-DOS
(D) SOFTWARE: PatentIn Release #1.0, Version #1.30
- (vi) CURRENT APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/851,089
(B) FILING DATE: 05-MAY-1997
(C) CLASSIFICATION:
- (vii) PRIOR APPLICATION DATA:
(A) APPLICATION NUMBER: US 08/835,292
(B) FILING DATE: 07-APR-1997
- (viii) ATTORNEY/AGENT INFORMATION:
(A) NAME: Elmore, Carolyn S.
(B) REGISTRATION NUMBER: 37,567
(C) REFERENCE/DOCKET NUMBER: EBC97-06A
- (ix) TELECOMMUNICATION INFORMATION:
(A) TELEPHONE: (781) 861-6240
(B) TELEFAX: (781) 861-9540

(2) INFORMATION FOR SEQ ID NO:1:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1362 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ix) FEATURE:

- (A) NAME/KEY: CDS
(B) LOCATION: 1..1359

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:1:

ATG ACC GAT CCA CGT CAG CTG CAC CTG GCC GGA TTC TTC TGT GCC GGC	48
Met Thr Asp Pro Arg Gln Leu His Leu Ala Gly Phe Phe Cys Ala Gly	
1 5 10 15	
AAC GTC ACG CAC GCC CAC GGA GCG TGG CGC CAC GCC GAC GAC TCC AAC	96
Asn Val Thr His Ala His Gly Ala Trp Arg His Ala Asp Asp Ser Asn	
20 25 30	
GGC TTC CTC ACC AAG GAG TAC TAC CAG CAG ATT GCC CGC ACG CTC GAG	144
Gly Phe Leu Thr Lys Glu Tyr Tyr Gln Gln Ile Ala Arg Thr Leu Glu	
35 40 45	
CGC GGC AAG TTC GAC CTG CTG TTC CTT CCC GAC GCG CTC GCC GTG TGG	192
Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Ala Leu Ala Val Trp	
50 55 60	
GAC AGC TAC GGC GAC AAT CTG GAG ACC GGT CTG CGG TAT GGC GGG CAA	240
Asp Ser Tyr Gly Asp Asn Leu Glu Thr Gly Leu Arg Tyr Gly Gln	
65 70 75 80	
GGC GCG GTG ATG CTG GAG CCC GGC GTA GTT ATC GCC GCG ATG GCC TCG	288
Gly Ala Val Met Leu Glu Pro Gly Val Val Ile Ala Ala Met Ala Ser	
85 90 95	
GTG ACC GAA CAT CTG GGG CTG GGC GCC ACC ATT TCC ACC ACC TAC TAC	336
Val Thr Glu His Leu Gly Leu Gly Ala Thr Ile Ser Thr Thr Tyr Tyr	
100 105 110	
CCG CCC TAC CAT GTA GCC CGG GTC GTC GCT TCG CTG GAC CAG CTG TCC	384
Pro Pro Tyr His Val Ala Arg Val Val Ala Ser Leu Asp Gln Leu Ser	
115 120 125	
TCC GGG CGA GTG TCG TGG AAC GTG GTC ACC TCG CTC AGC AAT GCA GAG	432
Ser Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Ser Asn Ala Glu	
130 135 140	
GCG CGC AAC TTC GGC TTC GAT GAA CAT CTC GAC CAC GAT GCC CGC TAC	480
Ala Arg Asn Phe Gly Phe Asp Glu His Leu Asp His Asp Ala Arg Tyr	
145 150 155 160	
GAT CGC GCC GAT GAA TTC CTC GAG GTC GTG CGC AAG CTC TGG AAC AGC	528
Asp Arg Ala Asp Glu Phe Leu Glu Val Val Arg Lys Leu Trp Asn Ser	
165 170 175	

TGG GAT CGC GAT GCG CTG ACA CTC GAC AAG GCA ACC GGC CAG TTC GCC	576
Trp Asp Arg Asp Ala Leu Thr Leu Asp Lys Ala Thr Gly Gln Phe Ala	
180 185 190	
GAT CCG GCT AAG GTG CGC TAC ATC GAC CAC CGC GGC GAA TGG CTC AAC	624
Asp Pro Ala Lys Val Arg Tyr Ile Asp His Arg Gly Glu Trp Leu Asn	
195 200 205	
GTA CGC GGG CCG CTT CAG GTG CCG CGC TCC CCC CAG GGC GAG CCT GTC	672
Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val	
210 215 220	
ATT CTG CAG GCC GGG CTT TCG GCG CGG GGC AAG CGC TTC GCC GGG CGC	720
Ile Leu Gln Ala Gly Leu Ser Ala Arg Gly Lys Arg Phe Ala Gly Arg	
225 230 235 240	
TGG GCG GAC GCG GTG TTC ACG ATT TCG CCC AAT CTG GAC ATC ATG CAG	768
Trp Ala Asp Ala Val Phe Thr Ile Ser Pro Asn Leu Asp Ile Met Gln	
245 250 255	
GCC ACG TAC CGC GAC ATA AAG GCG CAG GTC GAG GCC GCC GGA CGC GAT	816
Ala Thr Tyr Arg Asp Ile Lys Ala Gln Val Glu Ala Ala Gly Arg Asp	
260 265 270	
CCC GAG CAG GTC AAG GTG TTT GCC GCG GTG ATG CCG ATC CTC GGC GAG	864
Pro Glu Gln Val Lys Val Phe Ala Ala Val Met Pro Ile Leu Gly Glu	
275 280 285	
ACC GAG GCG ATC GCC AGG CAG CGT CTC GAA TAC ATA AAT TCG CTG GTG	912
Thr Glu Ala Ile Ala Arg Gln Arg Leu Glu Tyr Ile Asn Ser Leu Val	
290 295 300	
CAT CCC GAA GTC GGG CTT TCT ACG TTG TCC AGC CAT GTC GGG GTC AAC	960
His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Val Gly Val Asn	
305 310 315 320	
CTT GCC GAC TAT TCG CTC GAT ACC CCG CTG ACC GAG GTC CTG GGC GAT	1008
Leu Ala Asp Tyr Ser Leu Asp Thr Pro Leu Thr Glu Val Leu Gly Asp	
325 330 335	
CTC GCC CAG CGC AAC GTG CCC ACC CAA CTG GGC ATG TTC GCC AGG ATG	1056
Leu Ala Gln Arg Asn Val Pro Thr Gln Leu Gly Met Phe Ala Arg Met	
340 345 350	
TTG CAG GCC GAG ACG CTG ACC GTG GGA GAA ATG GGC CGG CGT TAT GGC	1104
Leu Gln Ala Glu Thr Leu Thr Val Gly Glu Met Gly Arg Arg Tyr Gly	
355 360 365	
GCC AAC GTG GGC TTC GTC CCG CAG TGG GCG GGA ACC CGC GAG CAG ATC	1152
Ala Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr Arg Glu Gln Ile	
370 375 380	

GCG GAC CTG ATC GAG ATC CAT TTC AAG GCC GGC GGC GCC GAT GGC TTC	1200
Ala Asp Leu Ile Glu Ile His Phe Lys Ala Gly Gly Ala Asp Gly Phe	
385 390 395 400	
ATC ATC TCG CCG GCG TTC CTG CCC GGA TCT TAC GAG GAA TTC GTC GAT	1248
Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Glu Glu Phe Val Asp	
405 410 415	
CAG GTG GTG CCC ATC CTG CAG CAC CGC GGA CTG TTC CGC ACT GAT TAC	1296
Gln Val Val Pro Ile Leu Gln His Arg Gly Leu Phe Arg Thr Asp Tyr	
420 425 430	
GAA GGC CGC ACC CTG CGC AGC CAT CTG GGA CTG CGT GAA CCC GCA TAC	1344
Glu Gly Arg Thr Leu Arg Ser His Leu Gly Leu Arg Glu Pro Ala Tyr	
435 440 445	
CTG GGA GAG TAC GCA TGA	1362
Leu Gly Glu Tyr Ala	
450	

(2) INFORMATION FOR SEQ ID NO:2:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 453 amino acids
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:2:

Met Thr Asp Pro Arg Gln Leu His Leu Ala Gly Phe Phe Cys Ala Gly	
1 5 10 15	
Asn Val Thr His Ala His Gly Ala Trp Arg His Ala Asp Asp Ser Asn	
20 25 30	
Gly Phe Leu Thr Lys Glu Tyr Tyr Gln Gln Ile Ala Arg Thr Leu Glu	
35 40 45	
Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Ala Leu Ala Val Trp	
50 55 60	
Asp Ser Tyr Gly Asp Asn Leu Glu Thr Gly Leu Arg Tyr Gly Gly Gln	
65 70 75 80	
Gly Ala Val Met Leu Glu Pro Gly Val Val Ile Ala Ala Met Ala Ser	
85 90 95	
Val Thr Glu His Leu Gly Leu Gly Ala Thr Ile Ser Thr Thr Tyr Tyr	
100 105 110	

Pro Pro Tyr His Val Ala Arg Val Val Ala Ser Leu Asp Gln Leu Ser
115 120 125

Ser Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Ser Asn Ala Glu
130 135 140

Ala Arg Asn Phe Gly Phe Asp Glu His Leu Asp His Asp Ala Arg Tyr
145 150 155 160

Asp Arg Ala Asp Glu Phe Leu Glu Val Val Arg Lys Leu Trp Asn Ser
165 170 175

Trp Asp Arg Asp Ala Leu Thr Leu Asp Lys Ala Thr Gly Gln Phe Ala
180 185 190

Asp Pro Ala Lys Val Arg Tyr Ile Asp His Arg Gly Glu Trp Leu Asn
195 200 205

Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val
210 215 220

Ile Leu Gln Ala Gly Leu Ser Ala Arg Gly Lys Arg Phe Ala Gly Arg
225 230 235 240

Trp Ala Asp Ala Val Phe Thr Ile Ser Pro Asn Leu Asp Ile Met Gln
245 250 255

Ala Thr Tyr Arg Asp Ile Lys Ala Gln Val Glu Ala Ala Gly Arg Asp
260 265 270

Pro Glu Gln Val Lys Val Phe Ala Ala Val Met Pro Ile Leu Gly Glu
275 280 285

Thr Glu Ala Ile Ala Arg Gln Arg Leu Glu Tyr Ile Asn Ser Leu Val
290 295 300

His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Val Gly Val Asn
305 310 315 320

Leu Ala Asp Tyr Ser Leu Asp Thr Pro Leu Thr Glu Val Leu Gly Asp
325 330 335

Leu Ala Gln Arg Asn Val Pro Thr Gln Leu Gly Met Phe Ala Arg Met
340 345 350

Leu Gln Ala Glu Thr Leu Thr Val Gly Glu Met Gly Arg Arg Tyr Gly
355 360 365

Ala Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr Arg Glu Gln Ile
370 375 380

Ala Asp Leu Ile Glu Ile His Phe Lys Ala Gly Gly Ala Asp Gly Phe
385 390 395 400

Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Glu Glu Phe Val Asp
405 410 415

Gln Val Val Pro Ile Leu Gln His Arg Gly Leu Phe Arg Thr Asp Tyr
420 425 430

Glu Gly Arg Thr Leu Arg Ser His Leu Gly Leu Arg Glu Pro Ala Tyr
435 440 445

Leu Gly Glu Tyr Ala
450

(2) INFORMATION FOR SEQ ID NO:3:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 1110 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:

- (A) NAME/KEY: CDS
- (B) LOCATION: 1..1107

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:3:

ATG ACG ACA GAC ATC CAC CCG GCG AGC GCC GCA TCG TCG CCG GCG GCG
Met Thr Thr Asp Ile His Pro Ala Ser Ala Ala Ser Ser Pro Ala Ala
1 5 10 15

CGC GCG ACG ATC ACC TAC AGC AAC TGC CCC GTG CCT AAT GCC CTG CTC
Arg Ala Thr Ile Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala Leu Leu
20 25 30

GCC GCG CTC GGC TCA GGT ATT CTG GAC AGT GCC GGG ATC ACA CTT GCC
Ala Ala Leu Gly Ser Gly Ile Leu Asp Ser Ala Gly Ile Thr Leu Ala
35 40 45

CTG CTG ACC GGA AAG CAG GGC GAG GTG CAC TTC ACC TAC GAC CGA GAT
Leu Leu Thr Gly Lys Gln Gly Glu Val His Phe Thr Tyr Asp Arg Asp
50 55 60

GAC TAC ACC CGC TTC GGC GGC GAG ATT CCG CCG CTG GTC AGC GAG GGA
Asp Tyr Thr Arg Phe Gly Gly Glu Ile Pro Pro Leu Val Ser Glu Gly
65 70 75 80

CTG CGT GCG CCG GGG CGG ACC CGC CTG CTG GGA CTG ACG CCG GTG CTG
Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Leu Thr Pro Val Leu
85 90 95

GGC CGC TGG GGC TAC TTC GTC CGG GGC GAC AGC GCG ATC CGC ACC CCG Gly Arg Trp Gly Tyr Phe Val Arg Gly Asp Ser Ala Ile Arg Thr Pro	336
100 105 110	
GCC GAT CTT GCC GGC CGC CGC GTC GGA GTA TCC GAT TCG GCC AGG AGG Ala Asp Leu Ala Gly Arg Arg Val Gly Val Ser Asp Ser Ala Arg Arg	384
115 120 125	
ATA TTG ACC GGA AGG CTG GGC GAC TAC CGC GAA CTT GAT CCC TGG CGG Ile Leu Thr Gly Arg Leu Gly Asp Tyr Arg Glu Leu Asp Pro Trp Arg	432
130 135 140	
CAG ACC CTG GTC GCG CTG GGG ACA TGG GAG GCG CGT GCC TTG CTG AGC Gln Thr Leu Val Ala Leu Gly Thr Trp Glu Ala Arg Ala Leu Leu Ser	480
145 150 155 160	
ACG CTC GAG ACG GCG GGG CTT GGC GTC GGC GAC GTC GAG CTG ACG CGC Thr Leu Glu Thr Ala Gly Leu Gly Val Gly Asp Val Glu Leu Thr Arg	528
165 170 175	
ATC GAG AAC CCG TTC GTC GAC GTG CCG ACC GAA CGA CTG CAT GCC GCC Ile Glu Asn Pro Phe Val Asp Val Pro Thr Glu Arg Leu His Ala Ala	576
180 185 190	
GGC TCG CTC AAA GGA ACC GAC CTG TTC CCC GAC GTG ACC AGC CAG CAG Gly Ser Leu Lys Gly Thr Asp Leu Phe Pro Asp Val Thr Ser Gln Gln	624
195 200 205	
GCC GCA GTC CTT GAG GAT GAG CGC GCC GAC GCC CTG TTC GCG TGG CTT Ala Ala Val Leu Glu Asp Glu Arg Ala Asp Ala Leu Phe Ala Trp Leu	672
210 215 220	
CCC TGG GCG GCC GAG CTC GAG ACC CGC ATC GGT GCA CGG CCG GTC CTA Pro Trp Ala Ala Glu Leu Glu Thr Arg Ile Gly Ala Arg Pro Val Leu	720
225 230 235 240	
GAC CTC AGC GCA GAC GAC CGC AAT GCC TAT GCG AGC ACC TGG ACG GTG Asp Leu Ser Ala Asp Asp Arg Asn Ala Tyr Ala Ser Thr Trp Thr Val	768
245 250 255	
AGC GCC GAG CTG GTG GAC CGG CAG CCC GAA CTG GTG CAG CGG CTC GTC Ser Ala Glu Leu Val Asp Arg Gln Pro Glu Leu Val Gln Arg Leu Val	816
260 265 270	
GAT GCC GTG GTG GAT GCA GGG CGG TGG GCC GAG GCC AAT GGC GAT GTC Asp Ala Val Val Asp Ala Gly Arg Trp Ala Glu Ala Asn Gly Asp Val	864
275 280 285	
GTC TCC CGC CTG CAC GCC GAT AAC CTC GGT GTC AGT CCC GAA AGC GTC Val Ser Arg Leu His Ala Asp Asn Leu Gly Val Ser Pro Glu Ser Val	912
290 295 300	

CGC CAG GGA TTC GGA GCC GAT TTT CAC CGC CGC CTG ACG CCG CGG CTC	960
Arg Gln Gly Phe Gly Ala Asp Phe His Arg Arg Leu Thr Pro Arg Leu	
305 310 315 320	
GAC AGC GAT GCT ATC GCC ATC CTG GAG CGT ACT CAG CGG TTC CTG AAG	1008
Asp Ser Asp Ala Ile Ala Ile Leu Glu Arg Thr Gln Arg Phe Leu Lys	
325 330 335	
GAT GCG AAC CTG ATC GAT CGG TCG TTG GCG CTC GAT CGG TGG GCT GCA	1056
Asp Ala Asn Leu Ile Asp Arg Ser Leu Ala Leu Asp Arg Trp Ala Ala	
340 345 350	
CCT GAA TTC CTC GAA CAA AGT CTC TCA CGC CAG GTC GAA GGG CAG ATA	1104
Pro Glu Phe Leu Glu Gln Ser Leu Ser Arg Gln Val Glu Gly Gln Ile	
355 360 365	
GCA TGA	1110
Ala	

(2) INFORMATION FOR SEQ ID NO:4:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 369 amino acids
 - (B) TYPE: amino acid
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:4:

Met Thr Thr Asp Ile His Pro Ala Ser Ala Ala Ser Ser Pro Ala Ala	
1 5 10 15	
Arg Ala Thr Ile Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala Leu Leu	
20 25 30	
Ala Ala Leu Gly Ser Gly Ile Leu Asp Ser Ala Gly Ile Thr Leu Ala	
35 40 45	
Leu Leu Thr Gly Lys Gln Gly Glu Val His Phe Thr Tyr Asp Arg Asp	
50 55 60	
Asp Tyr Thr Arg Phe Gly Gly Glu Ile Pro Pro Leu Val Ser Glu Gly	
65 70 75 80	
Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Leu Thr Pro Val Leu	
85 90 95	
Gly Arg Trp Gly Tyr Phe Val Arg Gly Asp Ser Ala Ile Arg Thr Pro	
100 105 110	

Ala Asp Leu Ala Gly Arg Arg Val Gly Val Ser Asp Ser Ala Arg Arg
115 120 125

Ile Leu Thr Gly Arg Leu Gly Asp Tyr Arg Glu Leu Asp Pro Trp Arg
130 135 140

Gln Thr Leu Val Ala Leu Gly Thr Trp Glu Ala Arg Ala Leu Leu Ser
145 150 155 160

Thr Leu Glu Thr Ala Gly Leu Gly Val Gly Asp Val Glu Leu Thr Arg
165 170 175

Ile Glu Asn Pro Phe Val Asp Val Pro Thr Glu Arg Leu His Ala Ala
180 185 190

Gly Ser Leu Lys Gly Thr Asp Leu Phe Pro Asp Val Thr Ser Gln Gln
195 200 205

Ala Ala Val Leu Glu Asp Glu Arg Ala Asp Ala Leu Phe Ala Trp Leu
210 215 220

Pro Trp Ala Ala Glu Leu Glu Thr Arg Ile Gly Ala Arg Pro Val Leu
225 230 235 240

Asp Leu Ser Ala Asp Asp Arg Asn Ala Tyr Ala Ser Thr Trp Thr Val
245 250 255

Ser Ala Glu Leu Val Asp Arg Gln Pro Glu Leu Val Gln Arg Leu Val
260 265 270

Asp Ala Val Val Asp Ala Gly Arg Trp Ala Glu Ala Asn Gly Asp Val
275 280 285

Val Ser Arg Leu His Ala Asp Asn Leu Gly Val Ser Pro Glu Ser Val
290 295 300

Arg Gln Gly Phe Gly Ala Asp Phe His Arg Arg Leu Thr Pro Arg Leu
305 310 315 320

Asp Ser Asp Ala Ile Ala Ile Leu Glu Arg Thr Gln Arg Phe Leu Lys
325 330 335

Asp Ala Asn Leu Ile Asp Arg Ser Leu Ala Leu Asp Arg Trp Ala Ala
340 345 350

Pro Glu Phe Leu Glu Gln Ser Leu Ser Arg Gln Val Glu Gly Gln Ile
355 360 365

Ala

(2) INFORMATION FOR SEQ ID NO:5:

(i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 1236 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(ix) FEATURE:
(A) NAME/KEY: CDS
(B) LOCATION: 1..1236

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:5:

ATG AAC GAA CTC GTC AAA GAT CTC GGC CTC AAT CGA TCC GAT CCG ATC	48
Met Asn Glu Leu Val Lys Asp Leu Gly Leu Asn Arg Ser Asp Pro Ile	
1 5 10 15	
GGC GCT GTG CGG CGA CTG GCC GCG CAG TGG GGG GCC ACC GCT GTT GAT	96
Gly Ala Val Arg Arg Leu Ala Ala Gln Trp Gly Ala Thr Ala Val Asp	
20 25 30	
CGG GAC CGG GCC GGC GGA TCG GCA ACC GCC GAA CTC GAT CAA CTG CGC	144
Arg Asp Arg Ala Gly Gly Ser Ala Thr Ala Glu Leu Asp Gln Leu Arg	
35 40 45	
GGC AGC GGC CTG CTC TCG CTG TCC ATT CCC GCC GCA TAT GGC GGC TGG	192
Gly Ser Gly Leu Leu Ser Leu Ser Ile Pro Ala Ala Tyr Gly Gly Trp	
50 55 60	
GGC GCC GAC TGG CCA ACG ACT CTG GAA GTT ATC CGC GAA GTC GCA ACG	240
Gly Ala Asp Trp Pro Thr Thr Leu Glu Val Ile Arg Glu Val Ala Thr	
65 70 75 80	
GTG GAC GGA TCG CTG GCG CAT CTA TTC GGC TAC CAC CTC GGC TGC GTA	288
Val Asp Gly Ser Leu Ala His Leu Phe Gly Tyr His Leu Gly Cys Val	
85 90 95	
CCG ATG ATC GAG CTG TTC GGC TCG GCG CCA CAA AAG GAA CGG CTG TAC	336
Pro Met Ile Glu Leu Phe Gly Ser Ala Pro Gln Lys Glu Arg Leu Tyr	
100 105 110	
CGC CAG ATC GCA AGC CAT GAT TGG CGG GTC GGG AAT GCG TCG AGC GAA	384
Arg Gln Ile Ala Ser His Asp Trp Arg Val Gly Asn Ala Ser Ser Glu	
115 120 125	
AAC AAC AGC CAC GTG CTC GAG TGG AAG CTT GCC ACC GCC GTC GAT	432
Asn Asn Ser His Val Leu Glu Trp Lys Leu Ala Ala Thr Ala Val Asp	
130 135 140	

GAT	GGC	GGG	TTC	GTC	CTC	AAC	GGC	GCG	AAG	CAC	TTC	TGC	AGC	GGC	GCC	480
Asp	Gly	Gly	Phe	Val	Leu	Asn	Gly	Ala	Lys	His	Phe	Cys	Ser	Gly	Ala	
145					150				155					160		
AAA	AGC	TCC	GAC	CTG	CTC	ATC	GTG	TTC	GGC	GTG	ATC	CAG	GAC	GAA	TCC	528
Lys	Ser	Ser	Asp	Leu	Leu	Ile	Val	Phe	Gly	Val	Ile	Gln	Asp	Glu	Ser	
					165				170					175		
CCC	CTG	CGC	GGC	GCG	ATC	ATC	ACC	GCG	GTC	ATT	CCC	ACC	GAC	CGG	GCC	576
Pro	Leu	Arg	Gly	Ala	Ile	Ile	Thr	Ala	Val	Ile	Pro	Thr	Asp	Arg	Ala	
					180				185					190		
GGT	GTT	CAG	ATC	AAT	GAC	GAC	TGG	CGC	GCA	ATC	GGG	ATG	CGC	CAG	ACC	624
Gly	Val	Gln	Ile	Asn	Asp	Asp	Trp	Arg	Ala	Ile	Gly	Met	Arg	Gln	Thr	
					195				200					205		
GAC	AGC	GGC	AGC	GCC	GAA	TTT	CGC	GAC	GTC	CGA	GTC	TAC	CCA	GAC	GAG	672
Asp	Ser	Gly	Ser	Ala	Glu	Phe	Arg	Asp	Val	Arg	Val	Tyr	Pro	Asp	Glu	
					210				215					220		
ATC	TTG	GGG	GCA	CCA	AAC	TCA	GTC	GTT	GAG	GCG	TTC	GTG	ACA	AGC	AAC	720
Ile	Leu	Gly	Ala	Pro	Asn	Ser	Val	Val	Glu	Ala	Phe	Val	Thr	Ser	Asn	
					225				230					235		240
CGC	GGC	AGC	CTG	TGG	ACG	CCG	GCG	ATT	CAG	TCG	ATC	TTC	TCG	AAC	GTT	768
Arg	Gly	Ser	Leu	Trp	Thr	Pro	Ala	Ile	Gln	Ser	Ile	Phe	Ser	Asn	Val	
					245				250					255		
TAT	CTG	GGG	CTC	GCG	CGT	GGC	GCG	CTC	GAG	GCG	GCA	GCG	GAT	TAC	ACC	816
Tyr	Leu	Gly	Leu	Ala	Arg	Gly	Ala	Leu	Glu	Ala	Ala	Ala	Asp	Tyr	Thr	
					260				265					270		
CGG	ACC	CAG	AGC	CGC	CCC	TGG	ACA	CCC	GCC	GGC	GTG	GCG	AAG	GCG	ACA	864
Arg	Thr	Gln	Ser	Arg	Pro	Trp	Thr	Pro	Ala	Gly	Val	Ala	Lys	Ala	Thr	
					275				280					285		
GAG	GAT	CCC	CAC	ATC	ATC	GCC	ACC	TAC	GGT	GAA	CTG	GCG	ATC	GCG	CTC	912
Glu	Asp	Pro	His	Ile	Ile	Ala	Thr	Tyr	Gly	Glu	Leu	Ala	Ile	Ala	Leu	
					290				295					300		
CAG	GGC	GCC	GAG	GCG	GCC	GCG	CGC	GAG	GTC	GCG	GCC	CTG	TTG	CAA	CAG	960
Gln	Gly	Ala	Glu	Ala	Ala	Ala	Arg	Glu	Val	Ala	Ala	Leu	Leu	Gln	Gln	
					305				310					315		320
GCG	TGG	GAC	AAG	GGC	GAT	GCG	GTG	ACG	CCC	GAA	GAG	CGC	GGC	CAG	CTG	1008
Ala	Trp	Asp	Lys	Gly	Asp	Ala	Val	Thr	Pro	Glu	Glu	Arg	Gly	Gln	Leu	
					325				330					335		
ATG	GTG	AAG	GTT	TCG	GGT	GTG	AAG	GCC	CTC	TCG	ACG	AAG	GCC	GCC	CTC	1056
Met	Val	Lys	Val	Ser	Gly	Val	Lys	Ala	Leu	Ser	Thr	Lys	Ala	Ala	Leu	
					340				345					350		

GAC ATC ACC AGC CGT ATT TTC GAG ACA ACG GGC TCG CGA TCG ACG CAT	1104
Asp Ile Thr Ser Arg Ile Phe Glu Thr Thr Gly Ser Arg Ser Thr His	
355 360 365	
CCC AGA TAC GGA TTC GAT CGG TTC TGG CGT AAC ATC CGG ACT CAT ACG	1152
Pro Arg Tyr Gly Phe Asp Arg Phe Trp Arg Asn Ile Arg Thr His Thr	
370 375 380	
CTG CAC GAT CCG GTA TCG TAT AAA ATC GTC GAT GTG GGG AAC TAC ACG	1200
Leu His Asp Pro Val Ser Tyr Lys Ile Val Asp Val Gly Asn Tyr Thr	
385 390 395 400	
CTC AAC GGG ACA TTC CCG GTT CCC GGA TTT ACG TCA	1236
Leu Asn Gly Thr Phe Pro Val Pro Gly Phe Thr Ser	
405 410	

(2) INFORMATION FOR SEQ ID NO:6:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 412 amino acids
- (B) TYPE: amino acid
- (C) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:6:

Met Asn Glu Leu Val Lys Asp Leu Gly Leu Asn Arg Ser Asp Pro Ile	
1 5 10 15	
Gly Ala Val Arg Arg Leu Ala Ala Gln Trp Gly Ala Thr Ala Val Asp	
20 25 30	
Arg Asp Arg Ala Gly Gly Ser Ala Thr Ala Glu Leu Asp Gln Leu Arg	
35 40 45	
Gly Ser Gly Leu Leu Ser Leu Ser Ile Pro Ala Ala Tyr Gly Gly Trp	
50 55 60	
Gly Ala Asp Trp Pro Thr Thr Leu Glu Val Ile Arg Glu Val Ala Thr	
65 70 75 80	
Val Asp Gly Ser Leu Ala His Leu Phe Gly Tyr His Leu Gly Cys Val	
85 90 95	
Pro Met Ile Glu Leu Phe Gly Ser Ala Pro Gln Lys Glu Arg Leu Tyr	
100 105 110	
Arg Gln Ile Ala Ser His Asp Trp Arg Val Gly Asn Ala Ser Ser Glu	
115 120 125	
Asn Asn Ser His Val Leu Glu Trp Lys Leu Ala Ala Thr Ala Val Asp	
130 135 140	

Asp Gly Gly Phe Val Leu Asn Gly Ala Lys His Phe Cys Ser Gly Ala
145 150 155 160

Lys Ser Ser Asp Leu Leu Ile Val Phe Gly Val Ile Gln Asp Glu Ser
165 170 175

Pro Leu Arg Gly Ala Ile Ile Thr Ala Val Ile Pro Thr Asp Arg Ala
180 185 190

Gly Val Gln Ile Asn Asp Asp Trp Arg Ala Ile Gly Met Arg Gln Thr
195 200 205

Asp Ser Gly Ser Ala Glu Phe Arg Asp Val Arg Val Tyr Pro Asp Glu
210 215 220

Ile Leu Gly Ala Pro Asn Ser Val Val Glu Ala Phe Val Thr Ser Asn
225 230 235 240

Arg Gly Ser Leu Trp Thr Pro Ala Ile Gln Ser Ile Phe Ser Asn Val
245 250 255

Tyr Leu Gly Leu Ala Arg Gly Ala Leu Glu Ala Ala Ala Asp Tyr Thr
260 265 270

Arg Thr Gln Ser Arg Pro Trp Thr Pro Ala Gly Val Ala Lys Ala Thr
275 280 285

Glu Asp Pro His Ile Ile Ala Thr Tyr Gly Glu Leu Ala Ile Ala Leu
290 295 300

Gln Gly Ala Glu Ala Ala Ala Arg Glu Val Ala Ala Leu Leu Gln Gln
305 310 315 320

Ala Trp Asp Lys Gly Asp Ala Val Thr Pro Glu Glu Arg Gly Gln Leu
325 330 335

Met Val Lys Val Ser Gly Val Lys Ala Leu Ser Thr Lys Ala Ala Leu
340 345 350

Asp Ile Thr Ser Arg Ile Phe Glu Thr Thr Gly Ser Arg Ser Thr His
355 360 365

Pro Arg Tyr Gly Phe Asp Arg Phe Trp Arg Asn Ile Arg Thr His Thr
370 375 380

Leu His Asp Pro Val Ser Tyr Lys Ile Val Asp Val Gly Asn Tyr Thr
385 390 395 400

Leu Asn Gly Thr Phe Pro Val Pro Gly Phe Thr Ser
405 410

(2) INFORMATION FOR SEQ ID NO:7:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 22 amino acids
(B) TYPE: amino acid
(C) STRANDEDNESS:
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:7:

Thr	Thr	Asp	Ile	His	Pro	Ala	Ser	Ala	Ala	Ser	Ser	Pro	Ala	Ala	Arg
1					5					10					15
Ala	Thr	Ile	Thr	Tyr	Ser										
					20										

(2) INFORMATION FOR SEQ ID NO:8:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 17 base pairs
(B) TYPE: nucleic acid
(C) STRANDEDNESS: single
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:8:

ACNGAYATHC AYCCNGC

17

(2) INFORMATION FOR SEQ ID NO:9:

- (i) SEQUENCE CHARACTERISTICS:
(A) LENGTH: 453 amino acids
(B) TYPE: amino acid
(D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:9:

Met	Thr	Gln	Gln	Arg	Gln	Met	His	Leu	Ala	Gly	Phe	Phe	Ser	Ala	Gly
1					5					10					15

57

Asn Val Thr His Ala His Gly Ala Trp Arg His Thr Asp Ala Ser Asn
20 25 30

Asp Phe Leu Ser Gly Lys Tyr Tyr Gln His Ile Ala Arg Thr Leu Glu
35 40 45

Arg Gly Lys Phe Asp Leu Leu Phe Leu Pro Asp Gly Leu Ala Val Glu
50 55 60

Asp Ser Tyr Gly Asp Asn Leu Asp Thr Gly Val Gly Leu Gly Gln
65 70 75 80

Gly Ala Val Ala Leu Glu Pro Ala Ser Val Val Ala Thr Met Ala Ala
85 90 95

Val Thr Glu His Leu Gly Leu Gly Ala Thr Ile Ser Ala Thr Tyr Tyr
100 105 110

Pro Pro Tyr His Val Ala Arg Val Phe Ala Thr Leu Asp Gln Leu Ser
115 120 125

Gly Gly Arg Val Ser Trp Asn Val Val Thr Ser Leu Asn Asp Ala Glu
130 135 140

Ala Arg Asn Phe Gly Ile Asn Gln His Leu Glu His Asp Ala Arg Tyr
145 150 155 160

Asp Arg Ala Asp Glu Phe Leu Glu Ala Val Lys Lys Leu Trp Asn Ser
165 170 175

Trp Asp Glu Asp Ala Leu Val Leu Asp Lys Ala Ala Gly Val Phe Ala
180 185 190

Asp Pro Ala Lys Val His Tyr Val Asp His His Gly Glu Trp Leu Asn
195 200 205

Val Arg Gly Pro Leu Gln Val Pro Arg Ser Pro Gln Gly Glu Pro Val
210 215 220

Ile Leu Gln Ala Gly Leu Ser Pro Arg Gly Arg Arg Phe Ala Gly Lys
225 230 235 240

Trp Ala Glu Ala Val Phe Ser Leu Ala Pro Asn Leu Glu Val Met Gln
245 250 255

Ala Thr Tyr Gln Gly Ile Lys Ala Glu Val Asp Ala Ala Gly Arg Asp
260 265 270

Pro Asp Gln Thr Lys Ile Phe Thr Ala Val Met Pro Val Leu Gly Glu
275 280 285

Ser Gln Ala Val Ala Gln Glu Arg Leu Glu Tyr Leu Asn Ser Leu Val
290 295 300

His Pro Glu Val Gly Leu Ser Thr Leu Ser Ser His Thr Gly Ile Asn
305 310 315 320

Leu Ala Ala Tyr Pro Leu Asp Thr Pro Ile Lys Asp Ile Leu Arg Asp
325 330 335

Leu Gln Asp Arg Asn Val Pro Thr Gln Leu His Met Phe Ala Ala Ala
340 345 350

Thr His Ser Glu Glu Leu Thr Leu Ala Glu Met Gly Arg Arg Tyr Gly
355 360 365

Thr Asn Val Gly Phe Val Pro Gln Trp Ala Gly Thr Gly Glu Gln Ile
370 375 380

Ala Asp Glu Leu Ile Arg His Phe Glu Gly Gly Ala Ala Asp Gly Phe
385 390 395 400

Ile Ile Ser Pro Ala Phe Leu Pro Gly Ser Tyr Asp Glu Phe Val Asp
405 410 415

Gln Val Val Pro Val Leu Gln Asp Arg Gly Tyr Phe Arg Thr Glu Tyr
420 425 430

Gln Gly Asn Thr Leu Arg Asp His Leu Gly Leu Arg Val Pro Gln Leu
435 440 445

Gln Gly Gln Pro Ser
450

(2) INFORMATION FOR SEQ ID NO:10:

- (i) SEQUENCE CHARACTERISTICS:
 - (A) LENGTH: 365 amino acids
 - (B) TYPE: amino acid
 - (C) TOPOLOGY: linear
- (ii) MOLECULE TYPE: protein
- (xi) SEQUENCE DESCRIPTION: SEQ ID NO:10:

Met Thr Ser Arg Val Asp Pro Ala Asn Pro Gly Ser Glu Leu Asp Ser
1 5 10 15

Ala Ile Arg Asp Thr Leu Thr Tyr Ser Asn Cys Pro Val Pro Asn Ala
20 25 30

Leu Leu Thr Ala Ser Glu Ser Gly Phe Leu Asp Ala Ala Gly Ile Glu
35 40 45

Leu Asp Val Leu Ser Gly Gln Gln Gly Thr Val His Phe Thr Tyr Asp
50 55 60

57

Gln Pro Ala Tyr Thr Arg Phe Gly Gly Glu Ile Pro Pro Leu Leu Ser
65 70 75 80

Glu Gly Leu Arg Ala Pro Gly Arg Thr Arg Leu Leu Gly Ile Thr Pro
85 90 95

Leu Leu Gly Arg Gln Gly Phe Phe Val Arg Asp Asp Ser Pro Ile Thr
100 105 110

Ala Ala Ala Asp Leu Ala Gly Arg Arg Ile Gly Val Ser Ala Ser Ala
115 120 125

Ile Arg Ile Leu Arg Gly Gln Leu Gly Asp Tyr Leu Glu Leu Asp Pro
130 135 140

Trp Arg Gln Thr Leu Val Ala Leu Gly Ser Trp Glu Ala Arg Ala Leu
145 150 155 160

Leu His Thr Leu Glu His Gly Glu Leu Gly Val Asp Asp Val Glu Leu
165 170 175

Val Pro Ile Ser Ser Pro Gly Val Asp Val Pro Ala Glu Gln Leu Glu
180 185 190

Glu Ser Ala Thr Val Lys Gly Ala Asp Leu Phe Pro Asp Val Ala Arg
195 200 205

Gly Gln Ala Ala Val Leu Ala Ser Gly Asp Val Asp Ala Leu Tyr Ser
210 215 220

Trp Leu Pro Trp Ala Gly Glu Leu Gln Ala Thr Gly Ala Arg Pro Val
225 230 235 240

Val Asp Leu Gly Leu Asp Glu Arg Asn Ala Tyr Ala Ser Val Trp Thr
245 250 255

Val Ser Ser Gly Leu Val Arg Gln Arg Pro Gly Leu Val Gln Arg Leu
260 265 270

Val Asp Ala Ala Val Asp Ala Gly Leu Trp Ala Arg Asp His Ser Asp
275 280 285

Ala Val Thr Ser Leu His Ala Ala Asn Leu Gly Val Ser Thr Gly Ala
290 295 300

Val Gly Gln Gly Phe Gly Ala Asp Phe Gln Gln Arg Leu Val Pro Arg
305 310 315 320

Leu Asp His Asp Ala Leu Ala Leu Leu Glu Arg Thr Gln Gln Phe Leu
325 330 335

Leu Thr Asn Asn Leu Leu Gln Glu Pro Val Ala Leu Asp Gln Trp Ala
340 345 350

Ala Pro Glu Phe Leu Asn Asn Ser Leu Asn Arg His Arg
355 360 365

(2) INFORMATION FOR SEQ ID NO:11:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 417 amino acids
- (B) TYPE: amino acid
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: protein

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:11:

Met Thr Leu Ser Pro Glu Lys Gln His Val Arg Pro Arg Asp Ala Ala
1 5 10 15

Asp Asn Asp Pro Val Ala Val Ala Arg Gly Leu Ala Glu Lys Trp Arg
20 25 30

Ala Thr Ala Val Glu Arg Asp Arg Ala Gly Gly Ser Ala Thr Ala Glu
35 40 45

Arg Glu Asp Leu Arg Ala Ser Ala Leu Leu Ser Leu Leu Val Pro Arg
50 55 60

Glu Tyr Gly Gly Trp Gly Ala Asp Trp Pro Thr Ala Ile Glu Val Val
65 70 75 80

Arg Glu Ile Ala Ala Asp Gly Ser Leu Gly His Leu Phe Gly Tyr
85 90 95

His Leu Thr Asn Ala Pro Met Ile Glu Leu Ile Gly Ser Gln Glu Gln
100 105 110

Glu Glu His Leu Tyr Thr Gln Ile Ala Gln Asn Asn Trp Trp Thr Gly
115 120 125

Asn Ala Ser Ser Glu Asn Asn Ser His Val Leu Asp Trp Lys Val Ser
130 135 140

Ala Thr Pro Thr Glu Asp Gly Gly Tyr Val Leu Asn Gly Thr Lys His
145 150 155 160

Phe Cys Ser Gly Ala Lys Gly Ser Asp Leu Leu Phe Val Phe Gly Val
165 170 175

Val Gln Asp Asp Ser Pro Gln Gln Gly Ala Ile Ile Ala Ala Ala Ile
180 185 190

Pro Thr Ser Arg Ala Gly Val Thr Pro Asn Asp Asp Trp Ala Ala Ile
195 200 205

Gly Met Arg Gln Thr Asp Ser Gly Ser Thr Asp Phe His Asn Val Lys
210 215 220

Val Glu Pro Asp Glu Val Leu Gly Ala Pro Asn Ala Phe Val Leu Ala
225 230 235 240

Phe Ile Gln Ser Glu Arg Gly Ser Leu Phe Ala Pro Ile Ala Gln Leu
245 250 255

Ile Phe Ala Asn Val Tyr Leu Gly Ile Ala His Gly Ala Leu Asp Ala
260 265 270

Ala Arg Glu Tyr Thr Arg Thr Gln Ala Arg Pro Trp Thr Pro Ala Gly
275 280 285

Ile Gln Gln Ala Thr Glu Asp Pro Tyr Thr Ile Arg Ser Tyr Gly Glu
290 295 300

Phe Thr Ile Ala Leu Gln Gly Ala Asp Ala Ala Ala Arg Glu Ala Ala
305 310 315 320

His Leu Leu Gln Thr Val Trp Asp Lys Gly Asp Ala Leu Thr Pro Glu
325 330 335

Asp Arg Gly Glu Leu Met Val Lys Val Ser Gly Val Lys Ala Leu Ala
340 345 350

Thr Asn Ala Ala Leu Asn Ile Ser Ser Gly Val Phe Glu Val Ile Gly
355 360 365

Ala Arg Gly Thr His Pro Arg Tyr Gly Phe Asp Arg Phe Trp Arg Asn
370 375 380

Val Arg Thr His Ser Leu His Asp Pro Val Ser Tyr Lys Ile Ala Asp
385 390 395 400

Val Gly Lys His Thr Leu Asn Gly Gln Tyr Pro Ile Pro Gly Phe Thr
405 410 415

Ser

(2) INFORMATION FOR SEQ ID NO:12:

- (i) SEQUENCE CHARACTERISTICS:
- (A) LENGTH: 4144 base pairs
 - (B) TYPE: nucleic acid
 - (C) STRANDEDNESS: single
 - (D) TOPOLOGY: linear
- (ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:12:

GGTCGAGAT CGATCTGACC GTCGAACCCG GCGCGTTCA AACCATCCTC TGGGCCTCT 60
TCTTGCACTT GACATAGGAA TCTCTACTAA ATAAATAGAT ATTTATTGCA CACTAAGTTC 120
GGTGATCAGG CCGACCGTGT GTCTCAAGTG CTCGCTCCGG GTGCCACGA GCTAAAGCGC 180
GCGATGCTGG GGCGACAGCG CTAGGCATTG CGTCCCTCA CACCAATGAT GAGATGATAC 240
GATGCGCATG ACCACTATCC GCACCTAGCA CGAAAGATCC GTGCATTTCG CGAATGCCAA 300
TGAAGAGGAC CGACGTACGG CAGCTTCCTA CGCTTCGCG CCATCGTTCA TAGCCAAGGT 360
CTTTTCGACG CCGGTTCGCG TGGCGACTG ACGGCGGTAG CGCCGCGACT ATTGTTCA 420
AACTCACGAG GATAAGAGCC TATGACCGAT CCACGTCAGC TGACCTGGC CGGATTCTTC 480
TGTGCCGGCA ACGTCACGCA CGCCCACGGA GCGTGGCGCC ACGCCGACGA CTCCAACGGC 540
TTCCTCACCA AGGAGTACTA CCAGCAGATT GCCCGCACGC TCGAGCGCGG CAAGTCGAC 600
CTGCTGTTCC TTCCCGACGC GCTCGCCGTG TGGGACAGCT ACGGCGACAA TCTGGAGACC 660
GGTCTGCGGT ATGGCGGGCA AGGCGCGGTG ATGCTGGAGC CGGGCGTAGT TATGCCGCG 720
ATGGCCTCGG TGACCGAAC A TCTGGGCTG GGCGCCACCA TTTCCACCAC CTACTACCCG 780
CCCTACCATG TAGCCCGGGT CGTCGCTTCG CTGGACCAGC TGTCCCTCCGG GCGAGTGTG 840
TGGAACGTGG TCACCTCGCT CAGCAATGCA GAGGCGCGA ACTTCGGCTT CGATGAACAT 900
CTCGACCACG ATGCCCGCTA CGATCGCGCC GATGAATTCC TCGAGGTCGT GCGCAAGCTC 960
TGGAACAGCT GGGATCGCGA TCGCTGACA CTCGACAAGG CAACCGGCCA GTTCGCCGAT 1020
CCGGCTAAGG TCGCTACAT CGACCACCGC GGCGAATGGC TCAACGTACG CGGGCCGCTT 1080
CAGGTGCCGC GCTCCCCCA GGGCGAGCCT GTCATTCTGC AGGCCGGCT TTCGGCGCGG 1140
GGCAAGCGCT TCGCCGGCG CTGGCGGGAC CGGGTGTCA CGATTCGCC CAATCTGGAC 1200
ATCATGCAGG CCACGTACCG CGACATAAAG GCGCAGGTCG AGGCCGCCGG ACGCGATCCC 1260
GAGCAGGTCA AGGTGTTTCG CGCGGTGATG CCGATCCTCG GCGAGACCGA GGCGATCGCC 1320
AGGCAGCGTC TCGAATACAT AAATTCGCTG GTGCATCCCG AAGTCGGCT TTCTACGTTG 1380
TCCAGCCATG TCGGGGTCAA CCTTGCCGAC TATTGCTCG ATACCCCGCT GACCGAGGTC 1440
CTGGCGATC TCGCCAGCG CAACGTGCC ACCCAACTGG GCATGTTGCG CAGGATGTTG 1500

CAGGCCGAGA CGCTGACCGT GGGAGAAATG GGCCGGCGTT ATGGCGCCAA CGTGGGCTTC 1560
GTCCCCCAGT GGGCGGGAAC CCGCGAGCAG ATCGCGGACC TGATCGAGAT CCATTTCAAG 1620
GCCGGCGGCG CCGATGGCTT CATCATCTCG CCGGCCTTCC TGCCCGGATC TTACGAGGAA 1680
TTCGTGATC AGGTGGTGCC CATCCTGCAG CACCGCGGAC TGTTCCGCAC TGATTACGAA 1740
GGCCGCACCC TGCGCAGCCA TCTGGGACTG CGTGAACCCG CATACTGGG AGAGTACGCA 1800
TGACGACAGA CATCCACCCG GCGAGCGCCG CATCGTCGCC GGCGGCGCGC GCGACGATCA 1860
CCTACAGCAA CTGCCCCGTG CCTAATGCC CGCTCGCCGC GCTCGGCTCA GGTATTCTGG 1920
ACAGTGCCGG GATCACACTT GCCCTGCTGA CCGGAAAGCA GGGCGAGGTG CACTTCACCT 1980
ACGACCGAGA TGACTACACC CGCTTCGGCG GCGAGATTCC GCCGCTGGTC AGCGAGGGAC 2040
TGGGTGCGCC GGGCGGACC CGCCTGCTGG GACTGACGCC GGTGCTGGC CGCTGGGCT 2100
ACTTCGTCCG GGGCGACAGC GCGATCCGCA CCCCCGGCCGA TCTTGCCGGC CGCCGCGTCG 2160
GAGTATCCGA TTCGGCCAGG AGGATATTGA CCGGAAGGCT GGGCGACTAC CGCGAACCTTG 2220
ATCCCTGGCG GCAGACCTG GTCGCGCTGG GGACATGGGA GGCGCGTGCC TTGCTGAGCA 2280
CGCTCGAGAC GGCGGGGCTT GGCGTCGGCG ACGTCGAGCT GACGCGCATC GAGAACCCGT 2340
TCGTCGACGT GCCGACCGAA CGACTGCATG CCGCCGGCTC GCTCAAAGGA ACCGACCTGT 2400
TCCCCGACGT GACCAGCCAG CAGGCCGAG TCCTTGAGGA TGAGCGGCC GACGCCCTGT 2460
TCGGTGGCT TCCCTGGCG CGCGAGCTCG AGACCCGCAT CGGTGCACGG CGGGTCCTAG 2520
ACCTCAGCGC AGACGACCGC AATGCCTATG CGAGCACCTG GACGGTGAGC GCCGAGCTGG 2580
TGGACCGGCA GCCCGAACTG GTGCAGCGC TCGTCGATGC CGTGGTGGAT GCAGGGCGGT 2640
GGGCCGAGGC CAATGGCGAT GTCGTCTCCC GCCTGCACGC CGATAACCTC GGTGTCAGTC 2700
CCGAAAGCGT CCGCCAGGGA TTCGGAGCCG ATTTTCACCG CCGCCTGACG CGCGGGCTCG 2760
ACAGCGATGC TATGCCATC CTGGAGCGTA CTCAGCGGTT CCTGAAGGAT GCGAACCTGA 2820
TCGATCGGTC GTTGGCGCTC GATCGGTGGG CTGCACCTGA ATTCCCTCGAA CAAAGTCTCT 2880
CACGCCAGGT CGAAGGGCAG ATAGCATGAA CGAACTCGTC AAAGATCTCG GCCTCAATCG 2940
ATCCGATCCG ATCGGGCGCTG TGCGGCGACT GGCGCGCAG TGGGGGGCCA CCGCTGTTGA 3000
TCGGGACCGG GCCGGCGGAT CGGCAACCCG CGAACTCGAT CAACTGCCG GCAGCGGCCT 3060
GCTCTCGCTG TCCATTCCCG CCGCATATGG CGGCTGGGC GCCGACTGGC CAACGACTCT 3120

GGAAAGTTATC CGCGAAGTCG CAACGGTGG A CGGATCGCTG GCGCATCTAT TCGGCTACCA	3180
CCTCGGCTGC GTACCGATGA TCGAGCTGTT CGGCTCGGCG CCACAAAAGG AACGGCTGTA	3240
CCGCCAGATC GCAAGCCATG ATTGGCGGGT CGGGAATGCG TCGAGCGAAA ACAACAGCCA	3300
CGTGCTCGAG TGGAAGCTTG CCGCCACCGC CGTCGATGAT GGCGGGTTCG TCCTCAACGG	3360
CGCGAAGCAC TTCTGCAGCG GCGCCAAAAG CTCCGACCTG CTCATCGTGT TCGGCGTGAT	3420
CCAGGACGAA TCCCCCTGC GCGGCGCGAT CATCACCGCG GTCATTCCA CCGACCGGGC	3480
CGGTGTTCA G ATCAATGACG ACTGGCGCC AATCGGGATG CGCCAGACCG ACAGCGGCAG	3540
CGCCGAATT CGCGACGTCC GAGTCTACCC AGACGAGATC TTGGGGGCAC CAAACTCAGT	3600
CGTTGAGGCG TTCGTGACAA GCAACCGCGG CAGCCTGTGG ACGCCGGCGA TTCAGTCGAT	3660
CTTCTCGAAC GTTTATCTGG GGCTCGCGCG TGGCGCGCTC GAGGCGGCAG CGGATTACAC	3720
CGGGACCCAG AGCCGCCCCT GGACACCCGC CGGCGTGGCG AAGGCGACAG AGGATCCCCA	3780
CATCATCGCC ACCTACGGTG AACTGGCGAT CGCGCTCCAG GGCGCCGAGG CGGCCGCGCG	3840
CGAGGTCGCG GCCCTGTTGC AACAGGCGTG GGACAAGGGC GATGCGGTGA CGCCCGAAGA	3900
GCGCGGCCAG CTGATGGTGA AGGTTTCGGG TGTGAAGGCC CTCTCGACGA AGGCCGCCCT	3960
CGACATCACC AGCCGTATTT TCGAGACAAC GGGCTCGCGA TCGACGCATC CCAGATAACGG	4020
ATTGATCGG TTCTGGCGTA ACATCCGGAC TCATACGCTG CACGATCCGG TATCGTATAA	4080
AATCGTCGAT GTGGGAACT ACACGCTCAA CGGGACATTC CCGGTTCCCG GATTTCAGTC	4140
ATGA	4144

(2) INFORMATION FOR SEQ ID NO:13:

(i) SEQUENCE CHARACTERISTICS:

- (A) LENGTH: 4144 base pairs
- (B) TYPE: nucleic acid
- (C) STRANDEDNESS: single
- (D) TOPOLOGY: linear

(ii) MOLECULE TYPE: DNA (genomic)

(xi) SEQUENCE DESCRIPTION: SEQ ID NO:13:

TCATGACGTA AATCCGGAA CGGGGAATGT CCCGTTGAGC GTGTAGTTCC CCACATCGAC	60
GATTTTATAC GATACCGGAT CGTGCAGCGT ATGAGTCCGG ATGTTACGCC AGAACCGATC	120

GAATCCGTAT CTGGGATGCG TCGATCGCGA GCCCGTTGTC TCGAAAATAC GGCTGGTGAT 180
GTCGAGGGCG GCCTTCGTCG AGAGGGCCTT CACACCCGAA ACCTTCACCA TCAGCTGGCC 240
GCGCTCTTCG GGCGTCACCG CATGCCCTT GTCCCACGCC TGTTGCAACA GGGCCGCGAC 300
CTCGCGCGCG GCCGCCTCGG CGCCCTGGAG CGCGATCGCC AGTTCACCGT AGGTGGCGAT 360
GATGTGGGGA TCCTCTGTCG CCTTCGCCAC GCCGGCGGGT GTCCAGGGC GGCTCTGGGT 420
CCGGGTGTA TCCGCTGCCG CCTCGAGCGC GCCACGCGCG AGCCCCAGAT AAACGTTCGA 480
GAAGATCGAC TGAATCGCCG GCGTCCACAG GCTGCCGCGG TTGCTTGTCA CGAACGCCTC 540
AACGACTGAG TTTGGTGCCC CCAAGATCTC GTCTGGTAG ACTCGGACGT CGCGAAATTG 600
GGCGCTGCCG CTGTCGGTCT GGCGCATCCC GATTGCGCGC CAGTCGTAT TGATCTGAAC 660
ACCGGCCCGG TCGGTGGAA TGACCGCGGT GATGATCGCG CCGCGCAGGG GGGATTGTC 720
CTGGATCACG CCGAACACGA TGAGCAGGTC GGAGCTTTG GCGCCGCTGC AGAAGTGCTT 780
CGCGCCGTTG AGGACGAACC CGCCATCATC GACGGCGGTG GCGGCAAGCT TCCACTCGAG 840
CACGTGGCTG TTGTTTCGC TCGACGCATT CCCGACCCGC CAATCATGGC TTGCGATCTG 900
GCGGTACAGC CGTCCCTTT GTGGCGCCGA GCCGAACAGC TCGATCATCG GTACGCAGCC 960
GAGGTGGTAG CCGAATAGAT GCGCCAGCGA TCCGTCCACC GTTGCAGCTT CGGGATAAAC 1020
TTCCAGAGTC GTTGGCCAGT CGGCGCCCCA GCCGCCATAT GCGGCGGGAA TGGACAGCGA 1080
GAGCAGGCCG CTGCCGCGCA GTTGCATCGAG TTCGGCGGTT GCCGATCCGC CGGCCCGGTC 1140
CCGATCAACA GCGGTGGCCC CCCACTGCGC GGCCAGTCGC CGCACAGCGC CGATCGGATC 1200
GGATCGATTG AGGCCGAGAT CTTTGACGAG TTCGTTCATG CTATCTGCC TTGACCTGG 1260
CGTGAGAGAC TTTGTTGAG GAATTCAAGGT GCAGCCCACC GATCGAGCGC CAACGACCGA 1320
TCGATCAGGT TCGCATCCTT CAGGAACCGC TGAGTACGCT CCAGGATGGC GATAGCATCG 1380
CTGTCGAGCC CGGGCGTCAG CGGGCGGTGA AAATCGGCTC CGAATCCCTG GCGGACGCTT 1440
TCGGGACTGA CACCGAGGTT ATCGGCGTGC AGGCCGGAGA CGACATCGCC ATTGGCCTCG 1500
GCCCACCGCC CTGCATCCAC CACGGCATCG ACGAGCCGCT GCACCAGTTC GGGCTGCCGG 1560
TCCACCAGCT CGGCGCTCAC CGTCCAGGTG CTCGCATAGG CATTGCGGTC GTCTGCGCTG 1620
AGGTCTAGGA CGGGCCGTGC ACCGATGCCG GTCTCGAGCT CGGCCGCCA GGGAAAGCCAC 1680
CGGAACAGGG CGTCGGCGCG CTCATCCTCA AGGACTGCCG CCTGCTGGCT GGTACGTCG 1740

GGGAACAGGT CGGTTCCCTT GAGCGAGCCG GCGGCATGCA GTCGTTGGT CGGCACGTCG	1800
ACGAACGGGT TCTCGATGCG CGTCAGCTCG ACGTCGCCGA CGCCAAGCCC CGCCGTCTCG	1860
AGCGTGCTCA GCAAGGCACG CGCCTCCCAT GTCCCCAGCG CGACCAGGGT CTGCCGCCAG	1920
GGATCAAGTT CGCGGTAGTC GCCCAGCCTT CCGGTCAATA TCCTCCTGGC CGAATCGGAT	1980
ACTCCGACGC GGCGGCCGGC AAGATCGGCC GGGGTGCGGA TCGCGCTGTC GCCCCGGACG	2040
AAGTAGCCCC AGCGGCCAG CACCGGCGTC AGTCCCAGCA GGCGGGTCCG CCCCGCGCA	2100
CGCAGTCCCT CGCTGACCAG CGCGGAATC TCGCCGCCGA AGCGGGTGT A GTCATCTCGG	2160
TCGTAGGTGA AGTGCACCTC GCCCTGCTTT CCGGTAGCA GGGCAAGTGT GATCCCGCA	2220
CTGTCCAGAA TACCTGAGCC GAGCGCGCG AGCAGGGCAT TAGGCACGGG GCAGTTGCTG	2280
TAGGTGATCG TCGCGCGCG CGCCGGCGAC GATGCGGCAC TCGCCGGTG GATGTCTGTC	2340
GTCATGCGTA CTCTCCCAGG TATGCGGGTT CACGCAGTCC CAGATGGCTG CGCAGGGTGC	2400
GGCCTTCGTA ATCAGTGCAG AACAGTCCGC GGTGCTGCAG GATGGGCACC ACCTGATCGA	2460
CGAATTCTC GTAAGATCCG GGCAGGAACG CCGCGAGAT GATGAAGCCA TCGCGCCGC	2520
CGGCCTTGAA ATGGATCTCG ATCAGGTCCG CGATCTGCTC GCGGGTTCCC GCCCACTGCG	2580
GGACGAAGCC CACGTTGGCG CCATAACGCC GGCCCATTTC TCCCACGGTC AGCGTCTCGG	2640
CCTGCAACAT CCTGGCGAAC ATGCCAGTT GGGTGGGCAC GTTGCCTGG GCGAGATCGC	2700
CCAGGACCTC GGTCAAGCGGG GTATCGAGCG AATAGTCGGC AAGGTTGACC CCGACATGGC	2760
TGGACAACGT AGAAAGCCCG ACTTCGGGAT GCACCAGCGA ATTATGTAT TCGAGACGCT	2820
GCCTGGCGAT CGCCTCGGTC TCGCCGAGGA TCGGCATCAC CGCGCAAAC ACCTTGACCT	2880
GCTCGGGATC GCGTCCGGCG GCCTCGACCT GCGCCTTAT GTCGCGGTAC GTGGCCTGCA	2940
TGATGTCCAG ATTGGCGAA ATCGTGAACA CCGCGTCCGC CCAGCGCCCG GCGAAGCGCT	3000
TGCCCCGCGC CGAAAGCCCG GCCTGCAGAA TGACAGGCTC GCCCTGGGG GAGCGCGCA	3060
CCTGAAGCGG CCCGCGTACG TTGAGCCATT CGCCCGGGTG GTCGATGTAG CGCACCTTAG	3120
CGGGATCGGC GAACTGGCCG GTTGCCTTGT CGAGTGTCA GCGATCGCGA TCCCAGCTGT	3180
TCCAGAGCTT GCGCACGACC TCGAGGAATT CATCGCGCG ATCGTAGCGG GCATCGTGGT	3240
CGAGATGTTA ATCGAAGCCG AAGTTGCGCG CCTCTGCATT GCTGAGCGAG GTGACCACGT	3300
TCCACGACAC TCGCCCGGAG GACAGCTGGT CCAGCGAAGC GACGACCCGG GCTACATGGT	3360

AGGGCGGGTA GTAGGTGGTG GAAATGGTGG CGCCCAGCCC CAGATGTTCG GTCACCGAGG	3420
CCATCGCGGC GATAACTACG CCGGGCTCCA GCATCACCGC GCCTTGCCCCG CCATACCGCA	3480
GACCGGTCTC CAGATTGTCG CCGTAGCTGT CCCACACGGC GAGCGCGTCG GGAAGGAACA	3540
GCAGGGTCGAA CTTGCCGCGC TCGAGCGTGC GGGCAATCTG CTGGTAGTAC TCCTTGGTGA	3600
GGAAGCCGTT GGAGTCGTCG GCGTGGCGCC ACGCTCCGTG GGCGTGCCTG ACGTTGCCGG	3660
CACAGAAGAA TCCGGCCAGG TGCAGCTGAC GTGGATCGGT CATAGGCTCT TATCCTCGTG	3720
AGTTTGAAAC GAATAGTCGC GGCCTACCG CCGTCAGTCG CCCACGCCAA CGGGCGTCGA	3780
AAAGACCTTG GCTATGAACG ATGGCGCGAA AGCGTAGGAA GCTGCCGTAC GTCGGTCCTC	3840
TTCATTGGCA TTTCGCAAAT GCACGGATCT TTCTGCTAG GTGCGGATAG TGGTCATGCG	3900
CATCGTATCA TCTCATCATT GGTGTGAGGG AACGCAATGC CTAGCGCTGT CGCCCCAGCA	3960
TCGCGCGCTT TAGCTCGTGG CAACCCGGAG CGAGCACTTG AGACACACGG TCGGCCTGAT	4020
CACCGAACCTT AGTGTGAAAT AAATATCTAT TTATTTAGTA GAGATTCCCTA TGTCAAGTGC	4080
AAGAAGAGGC CCCAGAGGAT GGTTTGAACC GCGCCGGTT CGACGGTCAG ATCGATCTCG	4140
AACC	4144